A Study of the Genetics of Diabetes Mellitus[1]

ARTHUR G. STEINBERG[2]

AND

RUSSELL M. WILDER[3]

Mayo Clinic, Rochester, Minnesota

THERE IS universal agreement that there is a large hereditary component in the causation of diabetes mellitus. In fact Joslin, in the "preamble" to Chapter 3 of the eighth edition of *The Treatment of Diabetes Mellitus*, makes the statement that "Heredity is the basis of diabetes." Disagreement arises, however, when an attempt is made to describe the genetic nature of this component. Further on in this same preamble Joslin indicates this uncertainty when he writes, "Would that experts in the genetic field would become interested in this problem."

The preamble is followed by a section written by White and Pincus in which they present an excellent review of the literature as well as of their analysis of data gathered in Joslin's clinic.

As illustrations of the disagreement which exists we may refer to the studies, involving sizable numbers of histories, of some workers who have come to quite different conclusions. Thus Cammidge (1928 and 1934) classified individual family pedigrees according to whether they could more readily be explained as resulting from a dominant or from a recessive gene. He concluded that the disease in those individuals with early onset and severe manifestation tended to be due to a recessive gene and that in those individuals with late onset and a mild expression, the disease tended to be due to a dominant gene (these genes were assumed to be at different loci). Hanhart (1950 and earlier), using the same method of analysis, maintained that all pedigrees may be explained by assuming a single recessive gene.

Pincus and White (1933) concluded from their analysis of the pedigrees of 523 diabetic and 153 nondiabetic patients that the data could be satisfactorily explained by assuming that the disease is due to a single recessive gene. However, Levit and Pessikova (1934) concluded from their analysis of 222 pedigrees

Received January 10, 1952.

[1] Work represented in this paper was carried out in the Division of Biometrics and Medical Statistics of the Mayo Clinic, Rochester, Minnesota.

[2] Geneticist Children's Cancer Research Foundation, Children's Medical Center, Boston 15, Massachusetts.

[3] Emeritus member, Division of Medicine, Mayo Clinic; now director, National Institute of Arthritis and Metabolic Diseases, National Institutes of Health, U. S. Public Health Service, Federal Security Administration.

that diabetes is due to a dominant gene with a penetrance of about 10 per cent and Harris (1950) has suggested, as a result of the analysis of data concerning the relatives of 1,241 diabetics, that "Many of the late-onset mild cases could be regarded as heterozygous for a gene which, in homozygous form, gives rise to the early-onset severe type of case." Penrose and Watson (1945) studied the sibs and parents of 442 patients with at least one affected relative and reported a sex-linked tendency in familial diabetes in a significant proportion of the families.

A priori we should expect that diabetes is genetically heterogeneous, as indeed many investigators have claimed. Such a situation is, of course, well known for many hereditary diseases in man as well as in other animals and plants. It also is well known that, despite genetic heterogeneity in some diseases, most of the cases are the result of one type of genetic change, as for example in albinism (Pipkin and Pipkin, 1942). It is possible that such is the situation with respect to diabetes mellitus.

THE MATERIAL

The material on which this report is based consists of information gathered over a two-year period by interviewing 1,981 consecutive patients at the Mayo Clinic who had the diagnosis of diabetes mellitus entered on their record for the first time during the current visit. Information concerning the patient's parents, sibs, spouse, and children was collected on a special form designed for this study.

We have no illusions concerning the accuracy of the information we have obtained. There are undoubtedly errors with regard to the ages of relatives, the presence or absence of diabetes, and the ages at onset. It is probable that there are inaccuracies with regard to the reporting of consanguinity between the patient's parents and between the patient and his spouse. In brief, we believe that every kind of error in reporting which could be made has been made in these data. However, we have exerted every effort to keep these errors to a minimum and, short of interviewing and examining every individual mentioned in each of these 1,981 pedigrees, we know of no way of obtaining more accurate data. We shall assume, as have all our predecessors, that the errors of reporting are random or at least are not sufficiently biased to disturb our conclusions.

THE DATA

Age at Onset.—The age at onset for those patients who were aware of their diabetes before they came to the clinic is the age reported by the patient. We do not know whether that age is the age at diagnosis or at detection of first symptoms. The possible error here is a compound one in that it involves the difficulty inherent in determining the age at onset plus the inaccuracy associ-

ated with an attempt to recall the time of an event which may have occurred long ago. This group composed 83 per cent (1,646/1,981) of the total sample. The remaining 17 per cent (335/1,981) did not know they had diabetes when they came to the clinic. The age at onset for this group of patients is reported as the age at diagnosis. This method of recording tends to increase the reported age at onset by an unknown margin depending on how long the disease may have been present and undiagnosed. The mean age at onset for the males in this group is 57 years and for the females 55 years, as contrasted with 46 and 45 years for the males and females in the group composed of those who knew of their diabetes prior to coming to the clinic.[4]

Table 1 shows the distribution of the ages of the 1,981 probands at onset. A Chi-square comparison of the age distributions of the male and female

TABLE 1. AGE OF PATIENT AT ONSET OF DIABETES

AGE, YEARS	TOTAL		MALES		FEMALES	
	Number	Per cent	Number	Per cent	Number	Per cent
0–9	51	2.57	23	2.02	28	3.33
10–19	107	5.40	59	5.18	48	5.70
20–29	134	6.76	81	7.11	53	6.29
30–39	230	11.61	125	10.98	105	12.47
40–49	455	22.98	267	23.44	188	22.33
50–59	603	30.44	340	29.85	263	31.24
60–69	336	16.96	201	17.65	135	16.03
70–79	60	3.03	39	3.42	21	2.49
80–89	5	0.25	4	0.35	1	0.12
Total.....	1,981	100.00	1,139	100.00	842	100.00
Average age at onset, years.....	47		48		47	

patients indicated that the distributions were not significantly different ($X^2_{[7]} = 8.048$; $P > 0.3$). The mean ages at onset were 48 years for the 1,139 males, 47 years for the 842 females, and 47 years for the combined total of 1,981 patients. We note here for future reference that the mean age at onset is essentially the same regardless of whether neither, one, or both of the patient's parents were diabetic, these ages being 47, 48 and 45 years respectively.

Table 2 presents a comparison between the present age of the patient and the age at onset for males and females separately. The correlation between these

[4] It is interesting to note that only 780 (47 per cent) of the 1,646 patients who were aware of their diabetes before coming to the clinic came because of their diabetes or its associated symptoms. Even more striking is the fact that only 59 (18 per cent) of the 335 whose diabetes was first diagnosed during their visit to the clinic came because overt symptoms of diabetes were present. None of the preceding values are influenced by the sex of the patient.

ages is high and essentially the same for both males and females, as is indicated by a comparison of the average present ages recorded in the extreme right columns for each decade of age at onset.

The relation between the age of the patient and of his affected parent at onset is presented in table 3. We present the data to make them available and shall not discuss them in detail because Harris (1950) has already shown that the apparent correlation is spurious and we have shown in two previous publications (Steinberg and Wilder, 1950 and in press) that the apparent "anticipation" is also spurious. There is nothing in the present data that contradicts either of these conclusions.

TABLE 2. AGE AT ONSET VERSUS PRESENT AGE OF PATIENT

AGE AT ONSET, YEARS	PRESENT AGE, YEARS									TOTAL	AVERAGE PRESENT AGE, YEARS
	0–9	10–19	20–29	30–39	40–49	50–59	60–69	70–79	80–89		
Males											
0–9	11	4	7	1						23	14
10–19		19	15	18	4	2	1			59	28
20–29			24	32	19	5	1			81	36
30–39				51	46	23	4	1		125	44
40–49					133	91	37	6		267	52
50–59						240	91	9		340	58
60–69							182	18	1	201	66
70–79								38	1	39	75
80–89									4	4	85
Total	11	23	46	102	202	361	316	72	6	1,139	
Females											
0–9	10	8	10							28	15
10–19		21	19	7		1				48	23
20–29			26	17	6	4				53	33
30–39				35	36	25	9			105	46
40–49					82	88	16	2		188	52
50–59						169	89	5		263	59
60–69							121	14		135	66
70–79								21		21	75
80–89									1	1	85
Total	10	29	55	59	124	287	235	42	1	842	

The relationship between the age of the patients and of their affected sibs at onset is shown in table 4. The correlation coefficient derived from these data is 0.549, which, while significantly lower than the value found by Harris (1950), 0.695 ($P < .01$), is significantly greater than zero. Harris showed that a large portion but not all of this high value of the correlation coefficient was due to the correlation which exists between the ages of sibs. Hence, most of the sibs

of the patients with early age of onset are still young and therefore could not have a late onset of diabetes. However, after correction for correlation due to this there remained a significant, but much lower, correlation between the age of the patient and of the affected sib at onset. The reality of this correlation may be illustrated in another more indirect way by reference to the data of

TABLE 3. RELATION BETWEEN AGE OF PATIENT AND AGE OF DIABETIC PARENT AT ONSET

AGE OF PATIENT AT ONSET, YEARS	AGE OF PARENT AT ONSET, YEARS								
	20–29	30–39	40–49	50–59	60–69	70–79	80–89	Total	Not stated
0–9			1					1	
10–19	1	1	1	2	3			8	
20–29	2	4	4	2	2	1		15	3
30–39		4	12	14	7	4	2	43	9
40–49		2	23	25	33	16	3	102	20
50–59		2	7	32	26	12	6	85	26
60–69		2	3	16	12	7	2	42	9
70–79				2	2	1		5	2
Total.....	3	15	51	93	85	41	13	301	69
Average age of patient, years...	22	39	42	50	49	48	51	48	49

TABLE 4. RELATION BETWEEN AGE OF PATIENT AND AGE OF DIABETIC SIB AT ONSET

AGE OF PATIENT AT ONSET, YEARS	AGE OF SIB AT ONSET, YEARS										
	0–9	10–19	20–29	30–39	40–49	50–59	60–69	70–79	80–89	Total	Not stated
0–9	1	4								5	1
10–19	2	4	1							7	0
20–29	3	6	7	7	3	5	1			32	5
30–39		1	7	10	18	7	1			44	12
40–49		2	4	16	37	33	17	2		111	25
50–59	1		9	10	26	59	18	5		128	32
60–69		2	3	5	13	30	26	5		84	12
70–79					1	5	5	1	1	13	1
Total.....	7	19	31	48	98	139	68	13	1	424	88
Average age of patient, years.	24	26	42	44	48	53	57	59	75	49	49

table 10 (which will be referred to again), in which it can be seen that the frequency of diabetes among the sibs of patients with early onset is as great as that among the sibs with late onset. This could arise only if a correlation existed between the age of patient and sib at onset.

While the existence of the correlation is clear its biological meaning is far

from being so. It is possible that a large portion of the correlation arises from peculiarities of establishing the presence of diabetes. For example, if diabetes is diagnosed in a child or young adult living at home, it is probable that others in the home will be examined for diabetes. Therefore, even mild diabetes, which might otherwise continue for a long time before diagnosis, would be detected soon after its onset. Among older persons who had already left home this would be less likely to occur. Other factors are the probable greater accuracy of knowledge of sibs when the patient is young than when the patient is old and the greater similarity of environment between the patients and sibs when the patient is young. The problem requires an intensive investigation of the

TABLE 5. BIRTH ORDER OF PATIENT VERSUS FAMILY SIZE

FAMILY SIZE	BIRTH ORDER														TOTAL
	1	2	3	4	5	6	7	8	9	10	11	12	13	16	
2	117	113													230
3	87	98	86												271
4	79	63	56	67											265
5	54	67	48	53	56										278
6	45	47	42	33	36	36									239
7	20	24	30	32	20	28	31								185
8	18	17	16	24	22	21	15	21							154
9	15	14	10	13	14	15	10	8	8						107
10	5	7	4	7	11	9	7	2	5	12					69
11	3	3	3	6	1	1	3	3	3	3	5				34
12	3	2	1	1	3	3	2	7	1	0	1	1			25
13	0	2	1	0	0	0	1	1	1	0	1	3	2		12
14	0	0	0	0	1	0	1	0	0	0	0	0	0		2
15	1	0	0	0	0	1	0	0	0	0	0	0	0		2
16	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
18	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Total observed...	447	457	297	236	164	114	71	42	18	15	7	4	2	1	1,875
Expected.....	438.1	438.1	323.1	232.7	166.5	110.9	71.0	44.6	25.3	13.4	6.5	3.4	1.3	0.1	

method of ascertaining the age at onset in the different families. It has characteristics in common with those encountered in "anticipation" and may very well have no greater biological significance than was shown to be the case for "anticipation" (Steinberg and Wilder, 1950 and in press).

Birth Order.—The data in table 5 show the birth order of the patients versus the number of children in the family who have survived the first year of life. (The 106 patients who had no siblings are not included in this table.) Only those who survived the first year of life are included because diabetes rarely occurs during the first year and therefore those not surviving this period are considered not to have been exposed to the risk of diabetes.

These data may be used to test the association between birth order and susceptibility to diabetes. If there is no association between susceptibility to diabetes and birth order, the expected number of patients in each birth order position for each family size is simply the number of families of the indicated size divided by that size. Thus there are 230 families of 2 children each; hence the number of patients expected for each birth order is 230/2 or 115.0. This process is repeated for each family size and then the contributions of each family size to birth order 1, birth order 2, 3, and so on, are summed. It is these sums which are entered on the bottom row of table 5. A X^2 comparison of the observed and expected values indicates that they are not significantly different ($X^2_{[11]} = 6.771$; $P \leqq 0.80$). It may be concluded that there is no correlation between birth order and susceptibility to diabetes.

Sex Ratios.—There were 1,139 male patients (57.5 per cent) and 842 female patients (42.5 per cent). The presence of an excess of males among diabetic patients, while not usual in the findings of other large studies, is not entirely

TABLE 6. SEX OF PATIENT VERSUS THAT OF AFFECTED PARENT OR SIB

PATIENT	AFFECTED RELATIVE				
	Parent*		Sib		
	Male	Female	Male	Female	Not stated
Male	89	114	147	121	2
Female	61	106	118	121	3
Total	150	220	265	242	5

* Only cases with 1 parent affected are included.

unique (Levit and Pessikova, 1934; Cammidge, 1934; Dahlberg and associates, 1947; and others), and furthermore is consistent with previous reports from this clinic (Wilder, Browne and Butt, 1940; Berkson, Gage and Wilder, 1947). Because the last paper cited contains a discussion of the possible reasons for this difference and because the subject will be dealt with in extenso elsewhere, it will not be discussed further here.

The relation between the sex of the patient and that of his affected parent or sib is shown in table 6. There is no indication of a significant association between the sex of the patient and that of the affected parent ($X^2_{[1]} = 2.033$; $P > 0.10$) nor between that of the patient and the affected sibs ($X^2_{[1]} = 1.510$; $P > 0.20$). However, we should like to call attention to the excess of affected mothers. The frequency of females among the affected parents (59.5 per cent) is significantly greater than that among the patients ($P < 0.001$). While we cannot account for this finding, we suggest that it may be due in part at least to a higher effective fertility among female diabetics as contrasted to male diabetics or to the patients' greater knowledge of their mothers than of their fathers or to a combination of these reasons.

Genetic Analysis.—Table 7 shows the number of affected offspring (patient and sibs) versus family size separated into three groups based on whether none, one or both parents were diabetic. From this table the average size of family and the frequency of diabetes among the patients' sibs may be computed for each of the three groups. The figures are presented in table 8. The large average

TABLE 7. NUMBER OF DIABETIC OFFSPRING (INCLUDING PATIENT) VERSUS FAMILY SIZE

FAMILY SIZE	NEITHER PARENT DIABETIC							ONE PARENT DIABETIC							BOTH PARENTS DIABETIC				
	Number of diabetic offspring							Number of diabetic offspring							Number of diabetic offspring				
	1	2	3	4	5	6	Total Fam.	1	2	3	4	5	7	Total Fam.	1	2	3	4	Total Fam.
1	91						91	13						13	2				2
2	178	7					185	40	2					42	3				3
3	201	19	1				221	39	6	4				49	1				1
4	190	21	6	2			219	35	6	3				44	2				2
5	192	16	8	1	1		218	36	12	9				57	2		1		3
6	152	34	4				190	33	11	3				47	1	1			2
7	113	31	1			1	146	14	13	4	3	2		36			1	2	3
8	95	20	4	1	1		121	17	9	5		1		32	1				1
9	65	7	3	1	1		77	15	7	3	1		1	27		2	1		3
10	34	17	4	1	1		57	6	3			1		10	1	1			2
11	13	10	4	1			28	2	2	1	1			6					
12	14	4	2				20	1		2	1	1		5					
13	4	6					10	1		1				2					
14	2						2												
15	1			1			2												
16	1						1												
18	1						1												
Total....	1,347	192	37	8	4	1	1,589	252	71	35	6	5	1	370	13	4	3	2	22

TABLE 8. AVERAGE FREQUENCY OF DIABETES AMONG SIBS OF PROBAND AND AVERAGE FAMILY SIZE VERSUS NUMBER OF DIABETIC PARENTS

NUMBER OF PARENTS DIABETIC	MEAN FAMILY SIZE	SIBS		
		Total	Diabetic	
			Number	Per cent
0	5.2	6,664	311	4.7
1	5.4	1,620	185	11.4
2	5.6	100	16	16.0

family size (5.2, 5.4, and 5.6 when none, one or both parents respectively are affected) is striking, particularly when it is recalled that only those individuals who survived the first year of life are included in the study. However, if it is noted that most of the families are completed (only 4 per cent of the patients were less than 20 years old and 1 per cent were less than 10 years of age [table

2]) and that most of them were completed more than a generation ago (only 17 per cent of the patients were less than 40 years of age [table 2]) and that a large proportion are rural families, the figures are not quite so surprising.

The average size of family does not appear to be reduced by the presence of diabetes among one or both parents, at least for those diabetic parents who have had one or more diabetic children. The same conclusion may be drawn from Pincus and White's and from Harris' data. In Pincus and White's sample the average number of offspring per family was 4.4 when neither parent was diabetic and 4.7 when one was (no figures are available for the 2 or 3 families with both parents affected) and in Harris' sample the values are 4.0, 4.4, and 5.4 when neither, one, or both parents respectively are diabetic. While in both sets of data the average family sizes are somewhat smaller than ours they show, as do ours, that the average size of family is not reduced by the presence of diabetes in one or both parents. Indeed, when all three sets of data are considered there is a suggestion that the average size of family may be increased when diabetes occurs in one or both of the parents. These remarks, naturally, are pertinent only to that group of diabetic parents who have produced at least one diabetic child and most probably are not correct as generalizations for all diabetics (Miller, 1946; Barns and Morgans, 1948).

The assumption of a single dominant gene with incomplete penetrance (Levit and Pessikova, 1934) is not consistent with the fact that in our sample, diabetes is more than twice as frequent among the sibs when one parent is diabetic than when neither parent is diabetic, nor is it consistent with the similar findings of Pincus and White (1933) and Harris (1950). Levit and Pessikova did not consider this point nor did they present their data in a manner which would permit others to consider it. If a dominant gene with incomplete penetrance were the correct explanation, the presence of overt diabetes in one parent would not be expected to increase the frequency of diabetes among the patients' sibs. Harris (1950) suggested that "Many of the late-onset mild cases could be regarded as heterozygous for a gene which, in homozygous form, gives rise to the early-onset severe type of case. The distribution of the homozygous and heterozygous cases, both in respect to age at onset and to severity, would be presumed to overlap to a certain extent. There would be incomplete manifestation, particularly of the heterozygotes. . . ." As is customary, Harris considers early onset as onset prior to age 30 years.

On the basis of this hypothesis, when onset is prior to age 30 years both of the patient's parents should be heterozygotes (we are assuming that none of the parents were homozygotes; see later) with the exception of the parents of an undefined number of patients who, although they became diabetic prior to age 30 years, are overlaps in the sense that they are heterozygotes with early onset. For the present purposes these overlaps may be ignored. When onset

in the patient is after age 30 years it is possible that both parents are heterozygotes, although in the usual case only one will be.

None of the parents in the 22 matings between affected individuals had onset prior to age 30 years and only 1 had onset prior to age 40 years. The 3 affected parents with onset prior to age 30 years (table 3) were all married to nondiabetic individuals. Hence if we disregard overlaps among the parents none of the matings involved homozygotes.

If we assume random mating the frequencies of the two types of realized matings (as already stated, we are considering the 3 parents with early onset as heterozygotes) which yield heterozygotes may be determined as follows: Let p equal the frequency of D, the gene leading to a susceptibility to diabetes, and $q = (1 - p)$ = frequency of d, its normal allele.

MATING	FREQUENCY	OFFSPRING		
		dd	*Dd*	*DD*
(1) $Dd \times dd$	$4\ pq^3$	$2\ pq^3$	$2\ pq^3$	
(2) $Dd \times Dd$	$4\ p^2q^2$	p^2q^2	$2\ p^2q^2$	p^2q^2

The relative frequency of heterozygotes from matings 1 and 2 equals q/p. With this ratio if we know the value of p we can compute the proportion of heterozygotes to be expected among the parents of patients who became diabetic after age 30 years. We do not know the value of p; however, we may safely assume that it does not exceed 0.5. If it equals 0.5 the ratio $q/p = 1$ and therefore equal numbers of heterozygotes are derived from matings 1 and 2 and consequently the proportion of heterozygotes among the parents of heterozygotes equals 0.75. If we assume a more reasonable value for p, say 0.10, the expected proportion of heterozygotes among the parents is 0.55. It follows therefore that the frequency of heterozygotes (affected parents) among the parents of patients with late onset of diabetes should be somewhat less than that among the parents of patients with early onset. The data are presented in table 9. In both sets of data (Harris' and the present sample) approximately twice as many diabetics occur among the parents of patients with late onset as among the parents of those with early onset.[5] This, of course, is exactly the reverse of the prediction.

If Harris is correct, the frequency of diabetes among the sibs of patients with onset prior to age 30 years should be greater than among the sibs of pa-

[5] Part of this excess is due to the fact that the parents of patients with late onset are older than those with early onset. However, the data of the present sample still show a higher frequency of diabetes among the parents of the patients with late onset after an approximate correction for this age difference based on the method used by Pincus and White (1933). The "corrected" values are 11.1 per cent for the parents of patients with early onset of diabetes and 17.5 per cent for those of patients with late onset of diabetes.

tients with onset after age 30 years because, as was already pointed out, most of the patients with early onset are homozygotes and therefore come from matings in which both parents are heterozygotes, while most patients with onset after 30 years are heterozygotes and are derived from matings which are mostly between a heterozygote and a normal. Furthermore, and very importantly, these frequencies should not be affected by the presence or absence of overt diabetes in one of the parents. The data are presented in table 10. Again we find that the data are not in agreement with the prediction. In both sets of data the frequency of diabetic sibs is the same, within statistical limits,

TABLE 9. FREQUENCY OF DIABETES AMONG PARENTS OF PATIENTS WITH EARLY AND LATE ONSET OF DIABETES

PATIENT'S AGE AT ONSET	PER CENT OF DIABETIC PARENTS	
	Harris' data	Present sample
Less than 30 years	3.3	5.0
30 years and more	6.2	11.4

TABLE 10. FREQUENCY OF DIABETES AMONG SIBS OF PATIENTS WITH EARLY AND LATE ONSET OF DIABETES

PATIENT'S AGE AT ONSET	ALL SIBS		NEITHER PARENT DIABETIC		ONE PARENT DIABETIC	
			Sibs		Sibs	
	Total	Per cent diabetic	Total	Per cent diabetic	Total	Per cent diabetic
Harris' data						
Less than 30 years	1,019	4.1	971	3.5	48	18.8
30 years and more	2,773	4.4	2,446	4.1	327	10.7
Present sample						
Less than 30 years	828	6.0	736	5.0	92	14.1
30 years and more	7,456	6.0	5,928	4.6	1,528	11.2

regardless of the age of the patient at the onset of diabetes, but in each age group the frequency of diabetic sibs is significantly greater when one parent is diabetic than when neither parent is diabetic.

Harris (1949) reported an excess of consanguinity among the parents of patients with early onset of diabetes but not among the parents of those with late onset. This excess was established by comparing his observed values with those found by Julia Bell (1940) in a survey of the hospital population of England. The over-all rate of consanguinity among the patients' parents which she found was 0.8 per cent. Among the diabetics in her sample she found 1.3

per cent with consanguineous parents, Harris found 1.5 per cent in his sample, and we find 1.8 per cent in our sample. This includes all degrees of consanguinity and patients of all ages. Table 11 shows a breakdown of our data. We have no base line with which to compare our data; however, we may note that there were no known marriages between first or second cousins among the parents of the patients with early onset and that the total rate of consanguinity was not higher among the parents of those with early onset than among those with late onset. The excess consanguinity observed by Harris occurred among the marriages of cousins of a degree higher than first cousins. In the present sample the frequency of consanguineous marriages of other than first degree among the parents of patients with diabetes of early onset is 1.4 per cent and for the parents of those with late onset 1.1 per cent. The comparable figures in Harris' sample are 1.2 and 0.3 respectively. We might examine the data differently by using Wright's (1922) coefficient of relation-

TABLE 11. CONSANGUINITY AMONG PARENTS

AGE OF PATIENT AT ONSET	CASES	DEGREE OF CONSANGUINITY									
		First cousins		Second cousins		Other cousins		Consanguinity unspecified		Total	
		Number	Per cent	Number	Per cent	Number	Per cent	Number	Per cent	Number	Per cent
Less than 30 years	292	0		0		2	0.7	2	0.7	4	1.4
30 years and more	1,689	12	0.7	7	0.4	8*	0.5	4	0.2	31	1.8
Total	1,981	12	0.6	7	0.4	10*	0.5	6	0.3	35	1.8

* One first cousin once removed.

ship, which computes the probability that an autosomal gene present in one individual will also be present in a specified relative by virtue of their common ancestry. The coefficients of relationship for the relationships pertinent to our discussion are as follows: Uncle-niece (one case in Harris' data) 1/4; first cousin, 1/8; first cousin once removed, 1/16; second cousin, 1/32 and third cousin, 1/128. To avoid fractions we shall set the coefficient for third cousins equal to 1 and change the others so that they retain their same relative values and then use these values to weight the observed frequencies of consanguineous marriages (all marriages not specified as to degree are considered third cousin marriages, as are all marriages more distant than those between second cousins), and calculate an average coefficient of relationship. The results are shown in table 12. The mean value is about ten times greater for the parents of patients with late onset than that for the parents of patients with early onset. In Harris' data the former value is about sixteen times as great as the latter (0.179 to 0.011). We may conclude that there is no very strong evidence

to indicate a greater frequency of consanguineous marriages among the parents of patients with early onset of diabetes as compared to that which occurs among the parents of those with late onset.

The possibility remains, however, that there is an increase in consanguinity among the parents of diabetics. This would be consistent with an hypothesis which assumes, as did Pincus and White (1933) and others, that diabetes is due to a single recessive gene. Another requirement of such an hypothesis is that the ratio of the frequencies of affected siblings derived from the three types of matings which yield recessive offspring (neither parent [$Aa \times Aa$], one parent [$Aa \times aa$], both parents [$aa \times aa$] affected) be as 1:2:4. This is true of conditions which are present congenitally. In a disease such as diabetes in which the age at onset is so variable the best we can expect is an approximation of these values, the closeness of the approximation depending in part on the age distributions of the offspring from each of the matings. In the present sample (table 8) these values are in the ratio of 1:2.4:3.4, which we

TABLE 12. AVERAGE RELATIVE VALUE OF COEFFICIENT OF RELATIONSHIP

AGE OF PATIENT AT ONSET	CASES	RELATIVE VALUE OF COEFFICIENT OF RELATIONSHIP	
		Total	Mean
Less than 30 years	292	4	0.014
30 years and more	1,689	239	0.141
Total	1,981	243	0.123

consider a reasonable approximation of the expected ratios. A third general requirement of this hypothesis is that the age of onset shall not be influenced by the presence or absence of diabetes in the parents. We have already noted (pp. 115 and 117) that the average ages at onset in the patients were essentially the same regardless of the presence or absence of diabetes in the parents.

While it is necessary that data fulfill these three conditions to fit a single-gene hypothesis, fulfilling them is not sufficient to prove that the hypothesis is in agreement with the data. An additional test is to derive numerical expectations and examine the closeness of the fit of these to the data. One approach, the one used by Pincus and White, is to correct for the variable age of onset of the disease and for the age distribution of the sibs; a second is to derive an expected frequency for each of the three types of mating which yield recessives and to compare the data with these values. We shall apply the latter method to our data and to those of other investigators.

This method of analysis assumes (1) that mating is random with respect to diabetes, (2) that all types of matings yielding diabetics are equally fertile and (3) that ascertainment is essentially equal for the matings that yield

diabetics. We have no evidence for the first assumption, pro or con, but it seems reasonable because most marriages occur before the onset of diabetes. The second assumption is known to be wrong when all diabetics are considered; however, as was shown in an earlier portion of the paper, fertility is equal among those matings which have yielded at least one diabetic offspring, therefore the assumption may be considered reasonable for our data. The third assumption offers no difficulty for the present sample because ascertainment is at a minimum.

On the basis of these assumptions if we assume that the diabetics are homozygous recessives and allow p to equal the frequency of the gene leading to diabetes and $q = (1 - p)$ to equal the frequency of its normal allele we can derive the proportions of all diabetics expected from each of the three types of matings, and from this we can estimate the value of p and test the fit of the observed data to the expected values. (Allan [1933] used a very similar

TABLE 13. EXPECTED FREQUENCIES OF MATINGS YIELDING DIABETICS, PROPORTION OF DIABETICS IN TOTAL POPULATION OF OFFSPRING AND PROPORTION OF DIABETICS ARISING FROM EACH MATING

MATING	1	2	$3 = 2/p^2$
	Frequency	Frequency of recessives in total population	Proportion of recessives arising from given mating
Neither parent diabetic	$4p^2q^2$	p^2q^2	q^2
One parent diabetic	$4p^3q$	$2p^3q$	$2pq$
Both parents diabetic	p^4	p^4	p^2

technic;[6] however he used a value of p obtained from an independent estimate of the frequency of diabetes in the population as a whole, and because this estimate yielded a high value of p, the deviations between the observed and expected values were considerable [$X^2_{(2)} = 7.064$, $P < 0.05$]. Consequently his cautious conclusion was, ". . . it seems possible that diabetes may be transmitted as a recessive unit character." We have analyzed his data by the method to be described presently and find that they do not differ significantly from the expected frequencies [table 14].) The details are shown in table 13. The maximum likelihood solution for p, using the values of column 3, is $p = \frac{b + 2c}{2N}$ where b equals the observed number of matings in which one parent was diabetic, c equals the observed number in which both parents were diabetic, and N equals the total number of matings. (This, of course, is the frequency of diabetes among the parents.) In the present sample $p = (370 + 2[22])/2(1{,}981) = 0.1045$; in Pincus and White's sample $p =$

[6] See also the theoretical discussion in Dahlberg and Hultkranz (1927).

$(80 + 2[3])/2(523) = 0.0822$; in Allan's sample $p = (17 + 2[2])/2(143) = 0.0734$, and in Harris' sample $p = (109 + 2[8])/2(1,241) = 0.0504$. These values of p may be used to compute the expected frequencies in each of the four samples of each of the three types of matings which yield recessives, namely, q^2, $2pq$, and p^2 when neither, one or both parents, respectively, are diabetic. The derived values and their comparison with the observed values are shown in table 14. The fit to our data, Pincus and White's and Allan's data is remarkably good, but Harris' data differ significantly from the expected values. We have no explanation of why Harris' sample is so different from the other three; however, Harris' sample was not a random one in that he selected for his purposes a disproportionately large number of cases with early onset of diabetes, and it may be because of this that the observed and theoretical values differ so greatly (see Steinberg and Wilder, in press, for further comment on Harris' sample).

TABLE 14. COMPARISON OF EXPECTED FREQUENCIES OF MATINGS YIELDING DIABETIC OFFSPRING WITH OBSERVED FREQUENCIES FOR FOUR SETS OF DATA

MATING	PRESENT SAMPLE		PINCUS AND WHITE		ALLAN		HARRIS	
	Expected	Observed	Expected	Observed	Expected	Observed	Expected	Observed
Both parents diabetic....	21.6	22	3.6	3	0.8	2	3.1	8
One parent diabetic.....	370.8	370	78.8	80	19.4	17	118.8	109
Neither parent diabetic..	1,588.6	1,589	440.6	440	122.8	124	1,119.1	1,124
$X^2_{(1)}$.....	0.009		0.119		2.109		8.573	
P.....	>0.90		>0.70		>0.10		<0.01	

DISCUSSION

On the basis of the preceding analysis it seems clear that neither the assumption of a dominant gene with incomplete penetrance nor the assumption that those persons with early onset of diabetes are homozygous for a gene which in the heterozygous condition leads to a late onset of diabetes is consistent with the data of our sample, or with those of the authors who made the original assumptions. The only single-gene hypothesis which is consistent with the several sets of the data is that a predisposition to diabetes is inherited as a simple recessive. Thus we confirm the conclusion advanced by Pincus and White in 1933. This, it seems to us, eliminates the necessity for assuming genetic heterogeneity of the type suggested by Cammidge (1934). We did not find, as did Penrose and Watson (1945), evidence of a sex-linked tendency in a significant proportion of the families; on the other hand, our data do not rule out the possibility that such a tendency may exist in some of the families.

We should like to emphasize that we do not believe that all cases of geneti-

cally determined diabetes are due to a simple recessive gene. Indeed, evidence exists in the literature (Burnstein and Patterson, 1949, as a recent example) and in our own files that in occasional pedigrees the disease is due to a dominant gene. Furthermore, it is not established that all cases due to a simple recessive gene are due to a change at the same locus. We do believe, however, that, as in albinism, the vast majority of the cases are due to a simple recessive mutation at one particular locus.

The method of analysis used in the previous section of the paper affords an estimate of the frequency of diabetes in the country as a whole if it is assumed that our sample is representative of the diabetics in the nation. This frequency (p^2) is 1.09 per cent of diagnosed diabetics in the present population. The estimate based on Pincus and White's sample is 0.68 per cent. The estimate based on their data is expected to be lower than the one based on ours because their sample was drawn approximately twenty years before ours was—that is, at a time when the population was relatively younger than it is now.

If a correction is made for the variable age at onset of diabetes and the age distribution of the population from which the sample was drawn, an estimate of the frequency of the gene in the population as a whole may be obtained. The corrected frequencies should be essentially the same for the two samples. Using the age distribution of the white population of the United States of the 1920 census to correct Pincus and White's sample, which was collected in 1932, and the population of the 1940 census to correct our data, which were collected over a two-year period beginning in May, 1949, and following the procedure used by Pincus and White, we obtain corrected values of $p = 0.22$ and 0.24 for the two samples respectively.

If these estimates of p are accepted as satisfactory we can predict that the total of diabetics (potential, undiagnosed and diagnosed) constitutes about 5 per cent of the population. In an earlier paragraph the frequency of diagnosed diabetics was estimated as about 1 per cent; hence there are approximately four times as many potential and undiagnosed diabetics in the population as there are diagnosed diabetics. As the average age of the population increases we may expect more diabetics to occur until the frequency of diabetics in the population approaches 5 per cent.

The estimates offered in preceding paragraphs reemphasize the importance to preventive medicine of gaining an understanding of the factors which cause a potential diabetic to become a frank diabetic.

The problem of explaining the variable age of onset of diabetes remains; in essence, this is the problem just referred to. Since most, if not all, diabetics are homozygous for the same recessive gene, the source of this variability must be sought at other loci and in the environment. If the correlation between the age of sibs at onset should prove to be of biologic significance, it would strongly suggest the presence of genetic modifiers affecting the time of

onset of the disease. A detailed longitudinal study of twins and of individuals with both parents diabetic could shed much light on this problem.

SUMMARY

1. The data concerning 1,981 consecutive diabetic patients and the information obtained from them concerning their parents and sibs were analyzed.

2. The modal age at onset was during the age interval 50–59 years for both male and female patients. The mean ages were 48 and 47 years respectively. The mean age at onset was not affected by the presence of diabetes in one or both parents.

3. No relation was found between birth order and susceptibility to diabetes.

4. No relation was found between the sex of the patient and that of his affected parent or sib. However, there is among the diabetic parents a significant excess of affected mothers.

5. The average size of family is not affected by the presence of diabetes in one or both parents, at least for those diabetic parents who have had one or more diabetic children.

6. The frequency in percentage of diabetes among the patients' sibs when diabetes is present in neither, one or both parents is 4.7, 11.4 and 16.0, respectively, in the ratio of 1:2.4:3.4. These figures are uncorrected for age.

7. The frequency of diabetes among the patients' parents is 10.45 per cent.

8. There is evidence of a possible increase of consanguinity among the parents of diabetics. The total rate of consanguinity was 1.8 per cent, and the frequency of marriages of first cousins was 0.6 per cent.

9. It is shown that only the hypothesis of a simple autosomal recessive gene is consistent with the data of this sample as well as those of some other published large samples.

10. It is estimated that the frequency of this gene lies between 20 and 25 per cent and that potential, diagnosed and undiagnosed diabetics constitute about 5 per cent of the population, while diagnosed diabetics form about 1 per cent of the population.

REFERENCES

ALLAN, W. 1933. Heredity in diabetes. *Ann. Int. M.* 6: 1272–1274.

BARNS, H. H. F. & MORGANS, M. E. 1948. Prediabetic pregnancy. *J. Obst. Gyn. Brit. Empire* 55: 449–554.

BELL, JULIA. 1940. A determination of the consanguinity rate in the general hospital population of England and Wales: A collaborative study summarized. *Ann. Eugen.* Cambr. 10: 370–391.

BERKSON, J., GAGE, R. P. & WILDER, R. M. 1947. Mortality and longevity among patients with diabetes mellitus. *Proc. Am. Diabetes Ass.* 7: 131–144.

BURNSTEIN, N. AND PATTERSON, M. 1949. Heredity in diabetes. Report of five generations of a diabetic family. *South. M. J.* 42: 119–120.

CAMMIDGE, P. J. 1928. Diabetes mellitus and heredity. *Brit. M. J.* 2: 738–741.

CAMMIDGE, P. J. 1934. Heredity as a factor in the Aetiology of Diabetes Mellitus. *Lancet.* Lond.. 1: 393–395.

DAHLBERG, G. & HULTKRANZ, J. V. 1927. Die Verbreitung eines monohybriden Erbmerkmals in einer Population und in der Verwandtschaft von Merkmalsträgern. Arch. *Rassenb.* 19: 129–165.

DAHLBERG, G., JORPES, E., KALLNER, S. & LICHTENSTEIN, A. 1947. Diabetes mellitus in Sweden: Statistical data on the number of diabetics, their state of health, working capacity, diet and treatment. *Acta med. scand.* Suppl. 188, 67 pp.

HANHART, E. 1950. Neue Forschungsergebnisse über die Vererbung des Diabetes mellitus sowie Anhaltspunkte für seine primare Genese im Stammhirn. *Arch. Julius Klaus-Stift.* 25: 586–598.

HARRIS, H. 1949. The incidence of parental consanguinity in diabetes mellitus. *Ann. Eugenics.* 14: 293–300.

HARRIS, H. 1950. The familial distribution of diabetes mellitus: A study of the relatives of 1,241 diabetic propositi. *Ann. Eugen.* Cambr. 15: 95–119.

JOSLIN, E. P., ROOT, H. F., WHITE, PRISCILLA, MARBLE, A. & BAILEY, C. C. 1946. The treatment of diabetes mellitus. Ed. 8, Philadelphia, Lea & Febiger, 861 pp.

LEVIT, S. G. & PESSIKOVA, L. N. 1934. The genetics of diabetes mellitus. *Proc. Maxim Gorky Med. Genet. Res. Inst. Moskva* 3: 132–147.

MILLER, H. C. 1946. The effect of diabetic and prediabetic pregnancies on the fetus and newborn infant. *J. Pediat.*, S. Louis 29: 455–461.

PENROSE, L. S. & WATSON, E. M. 1945. A sex-linked tendency in familial diabetes. *Proc. Am. Diabetes Ass.* 5: 163–177.

PINCUS, G. & WHITE, PRISCILLA. 1933. On the inheritance of diabetes mellitus. I. An analysis of 675 family histories. *Am. J. M. Sc.* 186: 1–14.

PIPKIN, A. C. & PIPKIN, S. B. 1942. Albinism in Negroes. *J. Hered.* 33: 419–427.

STEINBERG, A. G. & WILDER, R. M. 1950. A reconsideration of the phenomenon of anticipation in diabetes mellitus. *Proc. Staff Meet., Mayo Clin.* 25: 625–630.

STEINBERG, A. G. & WILDER, R. M. 1952. An analysis of the phenomenon of "anticipation" in diabetes mellitus. *Ann. Int. M.* 36: 1285–1296.

WILDER, R. M., BROWNE, H. C. & BUTT, H. R. 1940. Diseases of metabolism and nutrition: Review of certain recent contributions. *Arch. Int. M.* 65: 390–460.

WRIGHT, S. 1922. Coefficients of inbreeding and relationship. *Am. Natur.* 56: 330–338.

APPENDIX

TABLE A. AGE OF NONDIABETIC PARENTS: MATINGS IN WHICH BOTH WERE NONDIABETIC

AGE OF PATIENT AT ONSET, YEARS	AGE OF PARENT, YEARS*										
	10–19	20–29	30–39	40–49	50–59	60–69	70–79	80–89	90 and more	Not stated	Total
Mother											
0–9		4	18	17	8	3					50
10–19			9	33	31	20	3	2		1	99
20–29		2	5	12	41	36	16	3			115
30–39			10	12	33	57	45	12	2	1	172
40–49	1	6	17	21	35	74	102	53	10	6	325
50–59		5	28	40	50	106	121	113	17	5	485
60–69		4	13	18	23	50	76	87	11	3	285
70–79		1	4		3	8	14	17	6		53
80–89			1			2	1	1			5
Total.....	1	22	105	153	224	356	378	288	46	16	1,589
Father											
0–9		1	16	17	10	6					50
10–19			6	27	33	24	8	1			99
20–29		1	2	13	33	42	19	4		1	115
30–39		2	7	12	38	43	47	20	2	1	172
40–49		2	6	17	44	74	108	59	9	6	325
50–59		5	15	32	59	95	143	108	18	10	485
60–69		3	7	9	21	46	106	76	12	5	285
70–79			1	1	5	6	19	17	4		53
80–89						1	2	1	1		5
Total.....		14	60	128	243	337	452	286	46	23	1,589

* Present age if living, age at death if dead.

TABLE B. AGE OF NONDIABETIC PARENT: MATINGS IN WHICH ONE PARENT WAS DIABETIC

AGE OF PATIENT AT ONSET, YEARS	AGE OF PARENT, YEARS*									
	20–29	30–39	40–49	50–59	60–69	70–79	80–89	90 and more	Not stated	Total
Mother										
10–19			1	2	1					4
20–29				4	3	1	1	1		10
30–39		1		5	8	7	1			22
40–49		2	2	2	18	22	3	1		50
50–59		3	1	6	6	15	7	2		40
60–69		1	2	2	4	6	5	1		21
70–79				1				2		3
Total.....		7	6	22	40	51	17	7		150
Father										
0–9			1							1
10–19			1	1	2					4
20–29				2	4	2				8
30–39		2	1	5	9	8	4		1	30
40–49	1		3	11	22	22	11	2		72
50–59		3	1	13	19	22	13			71
60–69		1	3	4	7	4	9	1	1	30
70–79						3	1			4
Total.....	1	6	10	36	63	61	38	3	2	220

* Present age if living, age at death if dead.

TABLE C. AGE OF PARENTS AT ONSET OF DIABETES VERSUS AGE OF ONSET IN PATIENT: BOTH PARENTS DIABETIC

AGE OF PATIENT AT ONSET, YEARS	AGE OF PARENT AT ONSET, YEARS							
	30–39	40–49	50–59	60–69	70–79	80–89	Not stated	Total
Mother								
20–29		1						1
30–39	1		1	3	1			6
40–49		1	3	1	2		1	8
50–59		1	4		1		1	7
Total.....	1	3	8	4	4		2	22
Father								
20–29				1				1
30–39		1	1	2	1	1		6
40–49			2	1	1	1	3	8
50–59		1		3	1	1	1	7
Total.....		2	3	7	3	3	4	22

TABLE D. AGE OF ONSET OF DIABETES IN SIBS OF PATIENTS

AGE OF PATIENT AT ONSET, YEARS	AGE OF SIB AT ONSET, YEARS										
	0–9	10–19	20–29	30–39	40–49	50–59	60–69	70–79	80–89	Not stated	Total
Neither parent diabetic											
0–9	1	3								1	5
10–19	2	3	1								6
20–29	2	4	4	5	2	5				4	26
30–39		1	4	5	12	4	1			10	37
40–49		1	4	7	18	16	13	2		8	69
50–59	1		4	4	13	44	12	4		14	96
60–69		1	1	1	10	16	20	4		10	63
70–79						2	4	1	1	1	9
Total.	6	13	18	22	55	87	50	11	1	48	311
One parent diabetic											
0–9		1									1
10–19		1									1
20–29	1	2	3	2	1		1			1	11
30–39			3	4	5	1				2	15
40–49		1		9	18	16	4			14	62
50–59			5	6	8	14	5	1		18	57
60–69		1	2	4	3	14	6	1		2	33
70–79					1	3	1				5
Total.....	1	6	13	25	36	48	17	2		37	185
Both parents diabetic											
30–39				1	1	2					4
40–49					1	1				3	5
50–59					5	1	1				7
Total.....				1	7	4	1			3	16

TABLE E. PRESENT AGE* OF NONDIABETIC SIBS

AGE OF PATIENT AT ONSET, YEARS	AGE OF SIBS, YEARS											
	0–9	10–19	20–29	30–39	40–49	50–59	60–69	70–79	80–89	90 and more	Not stated	Total
Neither parent diabetic												
0–9	25	30	27	11	1							94
10–19	31	50	91	63	23	5	2	3				268
20–29	15	25	82	118	56	34	12					342
30–39	18	26	54	199	225	97	43	11	2			675
40–49	67	28	62	214	351	364	211	55	4			1,356
50–59	100	51	75	121	392	712	498	120	13		4	2,086
60–69	82	34	48	45	103	313	399	204	34	3	2	1,267
70–79	10	5	13	12	8	35	72	65	19		1	240
80–89		3		3		4	7	7	1			25
Total...	348	252	452	786	1,159	1,564	1,244	465	73	3	7	6,353
One parent diabetic												
0–9		1	2									3
10–19	1	2	7	8	3							21
20–29	2	3	7	21	16	3	3					55
30–39	13	5	16	49	64	46	14					207
40–49	25	16	23	91	144	118	51	13				481
50–59	27	7	19	36	89	168	96	23				465
60–69	11	3	3	4	16	74	53	23				187
70–79				1	1	2	5	7				16
Total...	79	37	77	210	333	411	222	66				1,435
Both parents diabetic												
20–29			1									1
30–39		1	7	8	8	8	2					34
40–49	1			2	20	8	1					32
50–59	1		1		7	6		2				17
Total...	2	1	9	10	35	22	3	2				84

* Present age if living, age at death if dead.